

Curator, Metafor and CMIP5, GO-ESSP, ESG, ...

GISS AR5 Workshop
New York, NY

V. Balaji

Princeton University

19 November 2008

Talk outline...

- 1 CMIP-5 timeline
 - Querying model characteristics
- 2 Curator and Metafor
 - Search
 - Gridspec
 - Analysis and Visualization
- 3 What we might be able to do for CMIP5

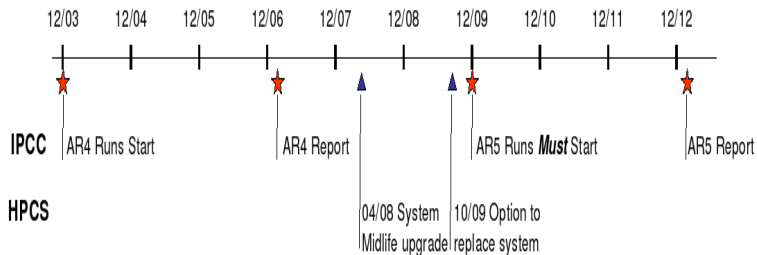
Talk outline ...

- 1 CMIP-5 timeline
 - Querying model characteristics

- 2 Curator and Metafor
 - Search
 - Gridspec
 - Analysis and Visualization

- 3 What we might be able to do for CMIP5

CMIP-5 timeline



- CMIP-5 more complicated than CMIP3 (AR4).
- CMIP-5 experiments will have links with other MIPs, principally PMIP and CFMIP, with their own timelines. Also CCMVal.
- A lot of people who participate in these projects also do ENSEMBLES (already begun...) may be a useful test case.
- CMOR-2 delivered for alpha-test in Oct 2008; beta-test in December 2008.

Points from GO-ESSP and WGCM meetings

- GO-ESSP meeting (Seattle, 17-19 September 2008) discussed a federation of **geographically distributed** data portals to maintain common conformant metadata for CMIP5.
- WGCM meeting (Paris 22-24 September 2008) developed consensus on the set of experiments for CMIP5/AR5. Includes “traditional” CMIP experiments; also includes a set of initialized runs for “decadal prediction” (think: ensembles, initial condition datasets. . .); carbon cycle experiments (new metadata); more detection and attribution runs (require description of forcings).
- discussion of filenaming conventions: **activity, institute, model, scenario/experiment, realm, data frequency, variable name, local ensemble member, version**
- broad acceptance of **native grids** (ocean group insists. . .)
- broad acceptance of need for structured info to replace AR4 questionnaire (SurveyMonkey looks great!)
- Controlled vocabulary for keys/values?

Can the database answer these questions?

- What's the difference between the NASA GISS-EH and GISS-ER models?
- Which runs from the GFDL CM2.1 model would I compare to isolate the effects of volcanoes on 20th century climate?
- Do volcano runs from GFDL CM2.1 and CCSM use the same forcing dataset?
- Which runs in the database include the **indirect effect of aerosols**?
- Retrieve “high cloud amount” from multiple models. (Not uniquely specified by standard name
`cloud_amount_in_atmospheric_layer!`)

Talk outline ...

- 1 CMIP-5 timeline
 - Querying model characteristics

- 2 Curator and Metafor
 - Search
 - Gridspec
 - Analysis and Visualization

- 3 What we might be able to do for CMIP5

ESG, Curator (ESC) and Metafor

- Earth System Grid (ESG) is tasked with building technology layers needed for linking a federation of geographically distributed data portals for CMIP5.
- Curator (US: 2005-2008) and Metafor (EU: 2008-2011) are proposal-driven projects to **organize climate model metadata**.
- Curator recognized that all the configuration information associated with a climate model could potentially be captured in the output metadata: **the workflow is the metadata**.
- Metafor continues this line of thought with a project to build a **Common Information Model (CIM)** for climate model data.
- All projects seek to work with existing standards (e.g CF) where available.

ESG/Curator Faceted Search

The screenshot shows a web browser window with the URL <http://cdp.ucar.edu:28080/query/queryESC.htm>. The browser's address bar and tabs are visible at the top. Below the browser window, the application header features a banner with the text "Earth System Grid" and navigation links: "Home", "Data", "About", and "Login". Below the banner, there are links for "Collection Browsing", "Simple Search", "Power Search (1)", "Power Search (2)", and "Data Visualization". The main content area is titled "CDP-Curator Search" and includes a "START OVER" link and a "TEXT SEARCH" input field. On the left, a "COMPONENT TYPE" list is displayed with the following items: Atmosphere, Atmospheric Chemistry, Atmospheric Dynamical Core, Atmospheric Dynamics, Atmospheric Physics, Biogeochemistry, Climate, Coastal Ocean, Coupled Atmosphere/Ocean General Circulation, Fisheries, General Circulation, Hydrology, Ice, Land, Land Ice, Magneto Hydro Dynamics, Ocean, Radiation, Sea Ice, Space Weather, Storm Surge, and Turbulence. The bottom of the browser window shows a "Done" status bar.

File Edit View Go Bookmarks Tools Help

<http://cdp.ucar.edu:28080/query/queryESC.htm>

Arts Books Canada Commercial DIY Film GFDL Google Libraries Mail Manuals Music News Photos Politics Princeton Science Sports Technology Tlemcen Weather

Google <http://...sg.ow/> WonderWe... Submit XML Proposed C... Metaedit Problem Ioa... Use Case Sc... Earth Syste... Google Cale... NCAR/UCAR... CDP-Curat...

Earth System Grid

Home Data About Login

Collection Browsing | Simple Search | Power Search (1) | Power Search (2) | Data Visualization

CDP-Curator Search

START OVER TEXT SEARCH

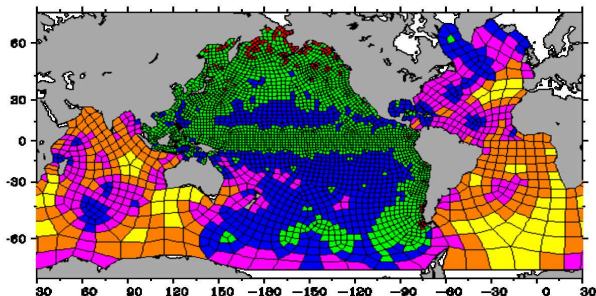
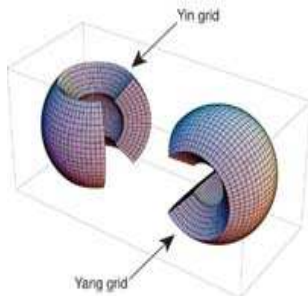
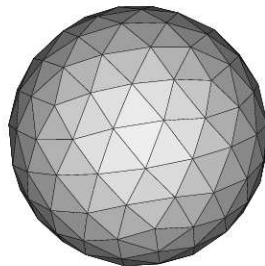
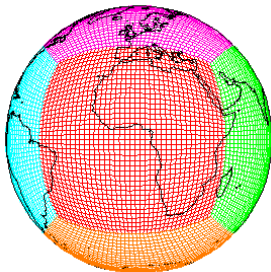
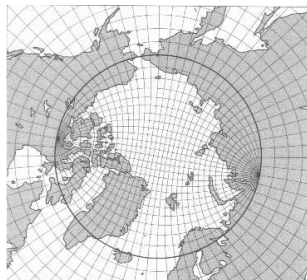
Select All Model Components Models Datasets Software Simulations

COMPONENT TYPE

- Atmosphere
- Atmospheric Chemistry
- Atmospheric Dynamical Core
- Atmospheric Dynamics
- Atmospheric Physics
- Biogeochemistry
- Climate
- Coastal Ocean
- Coupled Atmosphere/Ocean General Circulation
- Fisheries
- General Circulation
- Hydrology
- Ice
- Land
- Land Ice
- Magneto Hydro Dynamics
- Ocean
- Radiation
- Sea Ice
- Space Weather
- Storm Surge
- Turbulence

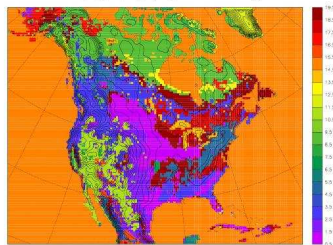
Done

Horizontal grids in use in ESMs

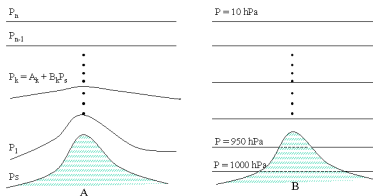


Vertical regridding: NARCCAP

GTOPO30 Topography (m) & GLCC Vegetation

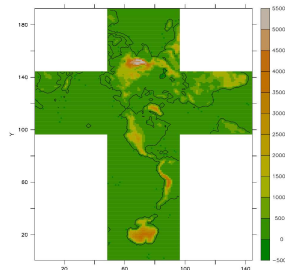
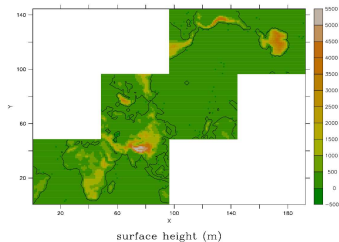


NX=155 NY=130 ds=50km CLAT=47.5 CLON=-97 Mercator



- The NARCCAP experiment is a MIP aimed at the “development of multiple high resolution regional climate scenarios for use in impacts assessments.”
- High-resolution models requires forcing data from global models and analysis in specified resolution, projection, and vertical levels.
- Data volumes are high: GFDL has chosen to supply data on its native grid (24 levels) instead of the required 40; in conjunction with a program for converting data from σ -hybrid to pressure.

Analysis and visualization



- `ferret`, a widely-used analysis and plotting utility is now capable of interpreting `gridspec` files and displaying the associated mosaic datasets. A “native” capability within `ferret` is being built, including as a “web service”!

http://www.gfdl.noaa.gov/~atw/ferret/cubed_sphere/

- Prototyping complex analyses (e.g “NINO3 SST spectra”) as web services.

Talk outline ...

- 1 CMIP-5 timeline
 - Querying model characteristics

- 2 Curator and Metafor
 - Search
 - Gridspec
 - Analysis and Visualization

- 3 What we might be able to do for CMIP5

Proposal for a Curator/Metafor product. . .

Small, self-contained, high-impact, within project scope. . .

- Begin with a CMIP5 AppCIM profile based on AR4 model metadata
- Add new fields (grid metadata; forcing fields and other datasets; in conjunction with PCMDI)
- Describe associated software?
- Telco on 10 October to discuss metadata changes for CMIP5 (includes Metafor, ESC, PCMDI, GFDL)
- Need to finalize the “CIM profile” for CMIP5 by December 2008?
- Create SurveyMonkey-based questionnaire from the above
- Release by Jan 2009?

Hypothetical CMIP5 case study

CMIP scientific question: response across a set of diverse models to "gregory-style forcing": response of a model at equilibrium to an abrupt increase in CO₂ to 4X.

- model "GFDL CM2.1"
- Scientific contact: Georgeanne at Princeton
- ocean "MOM4" uses tripolar grid, 1-deg in longitude, 1/3-1 deg variable in latitude: "OM3"
- atmos "FVLL" uses lat-lon grid 2x2.5 "M45"

Hypothetical CMIP5 case study (contd. . .)

CMIP scientific question: response across a set of diverse models to "gregory-style forcing": response of a model at equilibrium to an abrupt increase in CO₂ to 4X.

- model "GFDL CM2.1"
- simulation 1:
 - forcing dataset (boundary condition): constant CO₂ data from 2000, netCDF file interpolated to atm grid, "fvll-m45-co2-200-annual.nc"
 - run to equilibrium: 500 years in 100-year segments
 - save data from year 500 to year 700
 - data variables follow CMOR tables: static data, monthly mean, annual, seasonal, etc.
- simulation 2:
 - forcing dataset "fvll-m45-4Xco2-200-annual.nc"
 - restart model using initial condition from year 560 of simulation 1.
 - run for 200 years, save same data.

Hypothetical CMIP5 case study (contd. . .)

- model "HADCM3"
- Scientific contact: Hadrian at Exeter
- ocean "UM Ocean" uses lat-lon grid at 2x2 "HadOcn3"
- atmos "UM Atm" uses lat-lon grid 2x2.5 "HadAtm3"
- simulation 1:
 - forcing dataset (boundary condition): constant CO2 data from 2000
 - standard dataset processed by pre-processing software "HadPP"
run to equilibrium: 1000 years in 20-year segments
 - save data from year 1000 to year 1200
 - data variables follow CMOR tables: static data, monthly mean, annual, seasonal, etc.
- simulation 2:
 - model "HadCM3a", same as HadCM3, but code modified to multiply CO2 inputs by 4.
 - restart model using initial condition from year 1200 of simulation 1.
run for 200 years, save same data.

Potential pathways for analyzing multi-model ensembles

- Perform analyses “server-side” and download results. Likely to be restricted palette of analysis options; unlikely you would get to “upload” *your* analysis onto *their* servers.
- Download MME regridded onto target grid of your choice, organized with an ensemble axis (x, y, z, t, n) .
- Download native grid data *and* associated regridding tools (see NARCCAP example above) and construct common-grid ensemble yourself.